

AMENDMENT TO THE CLAIMS

1-26. (Canceled)

27. (Currently Amended) A method for evaluating a word segmentation language model, comprising:

building the word segmentation language model based on an annotated corpus;
applying the language model to a test corpus of unsegmented text different from the annotated corpus to provide an output indicative of words in the test corpus and a word type indication for each word, the word type indication being one of a plurality of word type indications;
comparing the word type indication for each word in the output of the language model with a predefined word type indicationssegmentation of words of the test corpus; and
evaluating the language model based on the comparison of the word type indication for each word in the output and the predefined word type indicationssegmentation to provide an indication of effectiveness of the language model as a function of the word type indications identified by the language model individual types of words.

28. (Currently Amended) The method of claim 27 wherein evaluating further comprises identifying words in the output that match words in the predefined word type indicationssegmentation.

29. (Currently Amended) The method of claim 27 wherein the word type indications ~~include~~comparing comprises comparing person names, location names, organization names, overlapping ambiguous strings and covering ambiguous strings in the output and the predefined word type indicationssegmentation.

30. (Currently Amended) The method of claim 29 wherein the indication of effectiveness is ~~calculated~~~~contacted~~ based on only the comparison of person names, location names, organization names, overlapping ambiguous strings and covering ambiguous strings.

31. (New) A method of evaluating word segmentation models, comprising:
using a first word segmentation model to segment a corpus of text into words and
apply tags to the words indicative of one of a plurality of word types, the
words and tags forming a first output;
using a second word segmentation model to segment the corpus of text into words
and apply tags to the words indicative of one of the plurality of word types,
the words and tags forming a second output;
comparing the first output to a predefined indication of words and tags of the
words indicative of one of the plurality of word types from the corpus of
text to provide a first set of values for each of the plurality of word types
indicative of how the first word segmentation model recognizes each of
the plurality of word types;
comparing the second output to the predefined indication of words and tags of the
words indicative of one of the plurality of word types from the corpus of
text to provide a second set of values for each of the plurality of word
types indicative of how the second word segmentation model recognizes
each of the plurality of word types; and
comparing the first set of values and the second set of values to determine
effectiveness of the first word segmentation model and the second word
segmentation model with respect to each of the plurality of word types.

32. (New) The method of claim 31 wherein the first set of values is based on matches between the first output and the predefined indication and wherein the second set of values is based on matches between the second output and the predefined indication.

33. (New) The method of claim 31 wherein the plurality of word types include person names, location names, organization names, overlapping ambiguous strings, and covering ambiguous strings.